Guidelines for Preparing Documents to be Distributed Via the World Wide Web

Rhode Island Department of Health
January 2000

Introduction

Distributing documents via the World Wide Web

Currently there is a move among the public and private sectors to provide a wide variety of information via the World Wide Web, or internet. This practice can allow a wider audience to be reached at a lower cost, and often in less time. The two most common formats used for this distribution are:

PDF - This format is preferable for documents with complex layouts or special fonts, and for documents which will primarily be downloaded and printed. PDF files embed special fonts and keep the look consistent when viewing and printing. These factors can be important for documents in which the layout is important, or when consistency is needed for accurate page numbering, etc. The Acrobat Reader software is required, but this is a free download from the web.

HTML - This format is excellent for providing information which is primarily textual. Control over formatting is limited, however, as the end-user can customize a browser to view web pages with settings to suit his or her tastes. Complex graphics and layouts are possible, but often labor-intensive. No additional software (besides a web browser) is required to view HTML.

If you are preparing a document which will reside on the web and are unsure of which format would work better, contact the Health Web Information Administrator and ask which format the document will be when posted to the website.

About this document

Publishing information on the web can be a very effective method for the distribution of information. However, the process has certain pitfalls which can make the move from traditional electronically-produced documents to web-distributable formats difficult. The Rhode Island Department of Health has provided information to the public via the internet for some time, but has encountered some of these obstacles during the process, reducing the speed and efficiency of the dissemination of information.

This document will explain some of the obstacles and how to avoid them. A few simple measures taken when preparing documents, keeping in mind the possible web distribution of the document, can make the move to the web faster and easier downstream. Each section pertains to a particular document source and final web format, with major points summarized at the end.

Preparing documents for PDF format

When using ANY program to generate documents for PDF

When using any program to create documents that will be converted to PDF format for distribution on the web, there area few issues which must be addressed in order to make the conversion to PDF relatively easy. The following steps should be followed before delivering documents to the Health Web Information Administrator for posting to a web site.

Use of Fonts

When a document is converted to a PDF file, the fonts used in the document are embedded into the file. This allows readers to view the document as it was created without having the same fonts installed. However, the original Word document cannot be viewed or converted properly (embedding the fonts) by the Health Web Information Administrator unless they have the required fonts installed on the system they are using to do the conversion.

For this reason, when delivering a document for conversion to PDF, include the files for any special fonts used in the document. These font files are located in the Windows directory, in a folder named "Fonts".
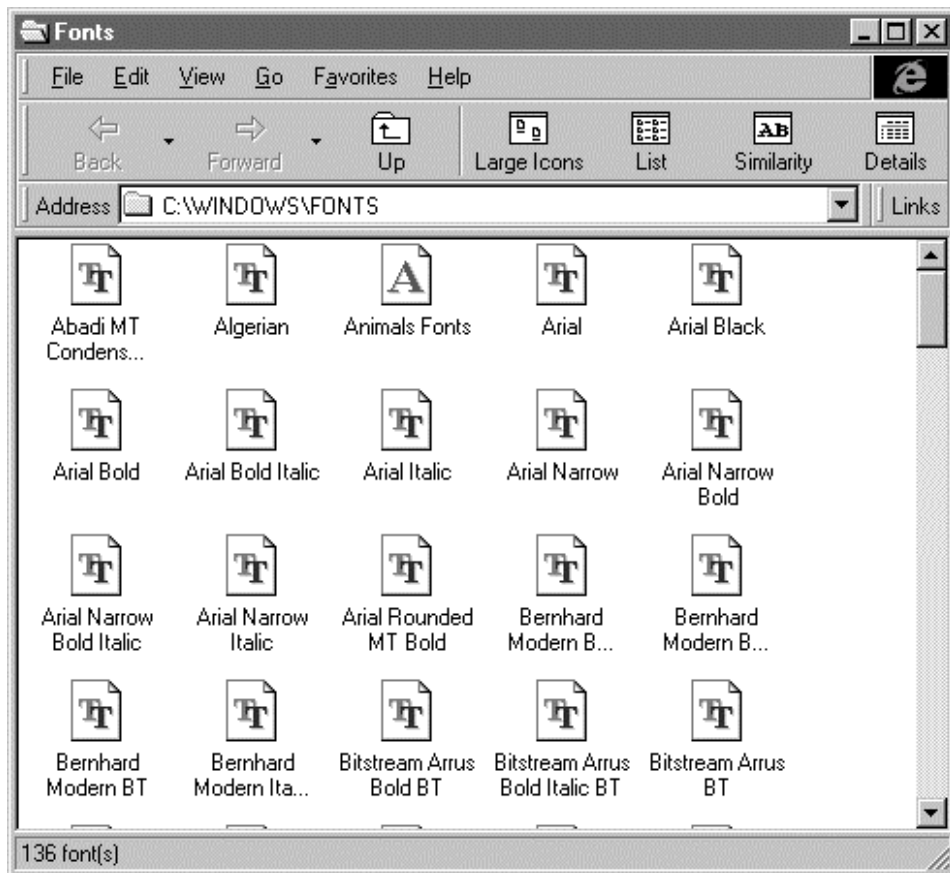
Also, when possible use True Type fonts in documents which will be converted to PDF format. This font type generally embeds into PDF files more readily. True Type fonts have the following icon:

Keep in mind that using too many different fonts in a document does not enhance the readability or appearance greatly. Fonts should be used sparingly, with special or exotic fonts used only for accents or titles. Standard fonts to use in documents for body text are Times/Times New Roman for a serif (has "feet") font and Arial or Helvetica for a san-serif (no "feet") font. Besides the font used for body text, no more than two fonts should be used for accents and headers. This keeps the document from becoming visually confusing, and reduces the number of fonts to be managed by the Health Web Information Administrator.

Margins

Sometimes when a document is converted to PDF, the margins in the original document are smaller than is allowed in the PDF conversion. The postscript drivers used by the system for the print-to-file operation in the PDF conversion will cut off content which is close to the edge of the page.

In order to avoid this problem, a minimum .5 inch margin should be used on all sides of the page. This includes headers and footers, which should also not be located closer than .5 inches from any edge.
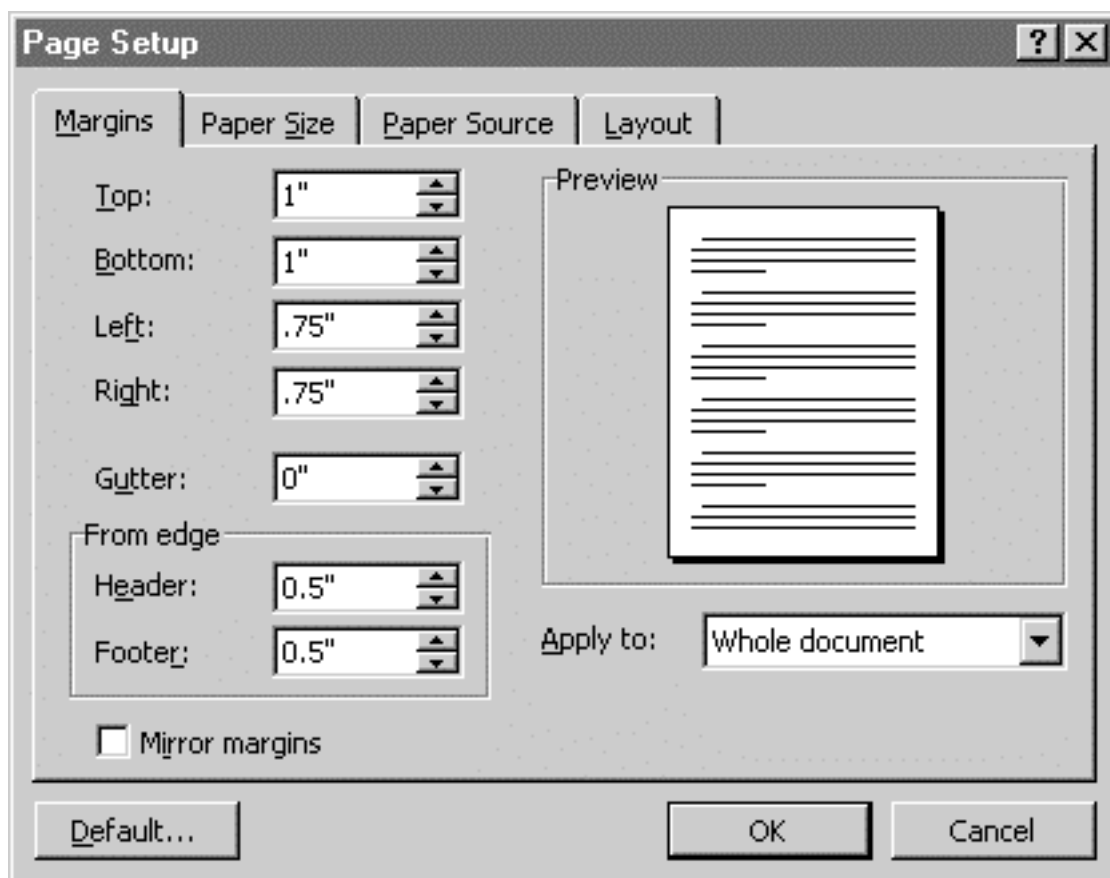
When using Microsoft Word to generate documents for PDF

In addition to the common issues involved with creating documents for PDF conversion, Microsoft Word files need special attention in certain areas. The following steps should be followed before delivering Word documents to the Health Web Information Administrator for posting to a web site.

Section Breaks and Margins

Microsoft Word allows a document to have sections, with a different set of page settings used in each. When using section breaks, care must be taken that the page settings do not change in a way which causes margins less than .5 inches, or a header or footer closer than .5 inches to a page edge. Sometimes this happens when a new section is started and settings return to a default.
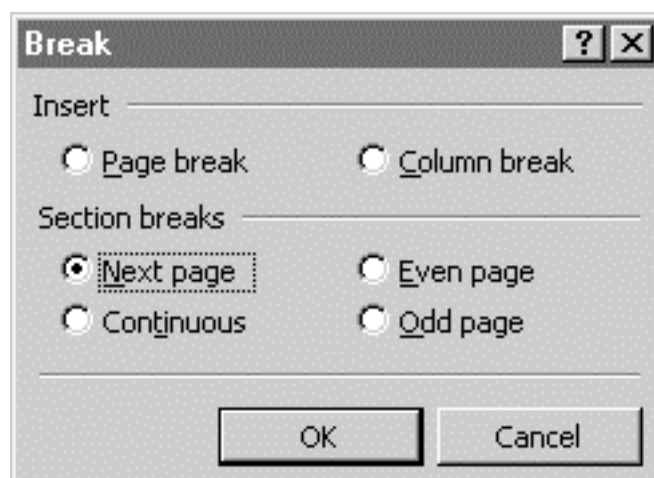
Section Breaks and Page Orientation

Section breaks allow Word users to change between Portrait and Landscape page orientations within the same document, but this can cause problems in PDF conversion. Sometimes the PDF creation software forces the document to convert as separate files, but does not allow the user to rename the files. This causes each new section to rewrite over the file during conversion, so a complete file is not produced.

If possible, keep all pages within a document in the same orientation. When a document MUST have different page orientations, provide the document in separate files for each orientation. This makes the conversion to PDF and file merge faster.

Page Numbering

If a document uses a table of contents referencing page numbers, they may not match if the document re-flows even slightly when opened on the computer used for the conversion to PDF. This problem can be avoided by inserting Page Breaks between pages, which forces the document to keep the page

numbering accurate. This also makes the problem easier to notice and repair by the person doing the conversion.

Headers and Footers

Sometimes a document will have the margins set correctly, but header or footer elements will be lost in the conversion to PDF. This usually means the "From Edge" settings are less than .5 inches in the Page Setup settings. Simply adjust these parameters to .5 inches or greater.

Keep in mind that the top and bottom margins must be set greater than the header and footer so that the page content does not flow into header and footer elements.

Text Boxes

Text boxes used in Microsoft Word often become distorted or display text incorrectly when converted to PDF. For this reason, text boxes should not be used in Word documents which will be distributed via the internet.

When using Microsoft Excel to generate documents for PDF

In addition to the common issues involved with creating documents for PDF conversion, page size should be taken into account when using Microsoft Excel.

- A common practice when using Excel is to generate spreadsheets in large format or with custom page sizes. Any spreadsheet created which will be converted to PDF should have text at least 8 points in size when scaled to print on Letter paper (8.5" x 11") size, or for large-format documents Tabloid (11" x 17") size. This avoids unreadable text when viewed or printed by the user, because users usually output at these sizes.

When using Microsoft Powerpoint to generate documents for PDF

The only major issues to be addressed about creating documents using Powerpoint for PDF conversion are covered in *When using ANY program to generate documents for PDF* shown above.

Preparing documents for HTML format

When using ANY program to generate documents for HTML

When using any program to create documents that will be converted to HTML format for distribution on the Web, there area few issues which must be addressed in order to make the conversion to HTML less problematic. The following steps should be followed before delivering documents to the Health Web Information Administrator for posting to a web site.

Use of Fonts

When preparing a document which will be posted to the web as an HTML page, the use of fonts must be kept in mind. Any special fonts used in the document must be installed on the visitor's system to display as intended when the document was created, and this is an unreliable assumption.
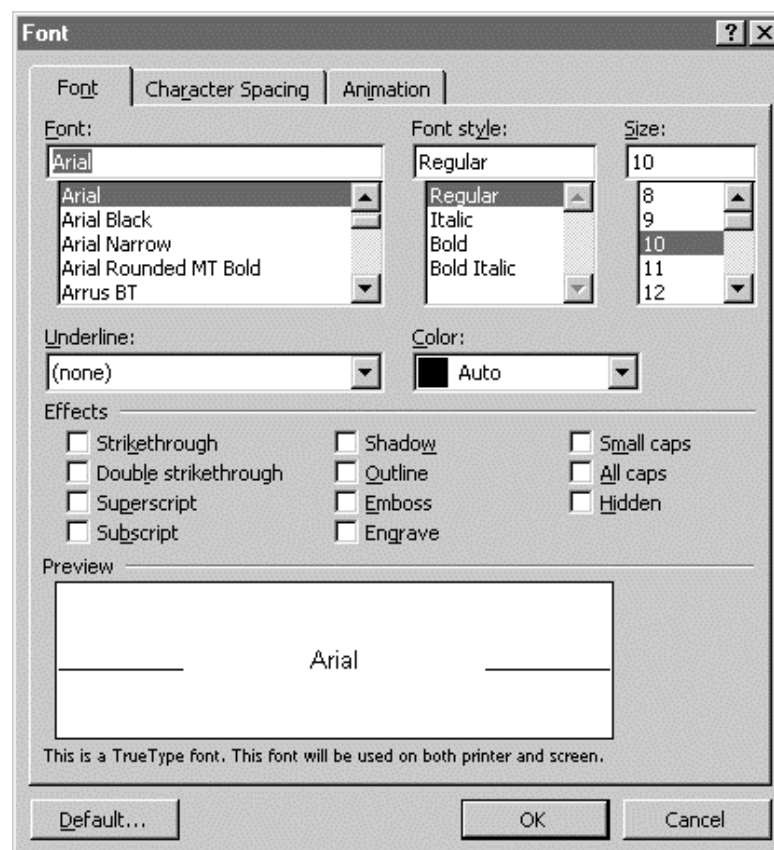
A few fonts are relatively safe for use in HTML documents, as these are common or default fonts on most systems:

Serif (has "feet")	Times or Times New Roman, often used in body text
San-Serif (no "feet")	Arial and Helvetica, often used for headlines and sub-heads

By using these fonts you can be sure that the document you create and the final web version will look similar. If a special font MUST be used, it can be provided as a graphic (see next entry) in order to display properly. This makes the text non-editable and the page will load slower, however.

Font Sizes

Most web browsers display text at preset sizes, which are usually limited to the following:

9 pt.   10pt.   12pt.   14pt.   18pt.   24pt.

If a different font size is used, many browsers will simply force the text into one of these sizes, which could change the look of the document. If these sizes are used when creating the document, it is probable that the text will display at the intended size on browsers with default settings.

Photos or Graphics

Images which will be included in a web page must be rasterized, or pixel-based, to display in a web browser. There are two graphic file formats which are supported by web browsers for display on web pages:

GIF   Good for images with large areas of solid color or straight lines. This format reduces file size by dithering the image into a smaller color palette, which can cause a jagged or noisy appearance on complex images. However, this format displays accurately on older systems using 256 colors if a "web palette" is specified when saving the image (if possible).

JPG   Good for photos or complex graphics. Uses a mathematical compression which allows images to appear in a greater number of colors.

When providing photos or graphics to be included in a web page, save the file as one of these formats before submitting it to the Health Web Information Administrator if possible. Also, if you know the size you would like the image to display at, try to save it at that scale before hand. Avoid very large graphics which may make the page download slowly, however.

When using Microsoft Word to generate documents for HTML

When using Microsoft Word to create documents that will be converted to HTML format for distribution on the Web, there is one issue that must be addressed in order to avoid problems.

Text Boxes

Text boxes often do not display properly after conversion to HTML. For this reason, text boxes should not be used in documents which will be distributed via the internet.

When using Microsoft Excel to generate documents for HTML

When using Microsoft Excel to generate documents which will reside on the web as HTML, the following should be kept in mind.

Scrolling

Many spreadsheets are large format or contain many columns. However, most users have monitors which have a 14" or 15" diagonal viewable area. This means that a wide document will force the user to scroll sideways, which can be annoying, as scrolling vertically is more common on computer screens. If possible, create spreadsheets which the user will view by scrolling vertically instead of horizontally.

When using layout programs (i.e. Visio, Pagemaker, Powerpoint) for HTML

Programs which allow complicated layouts incorporating graphics and text can create attractive presentations or ad material, but these documents do not easily convert to HTML. If material created with these programs is to be incorporated into a web page, the following points should be kept in mind.

Download speed

These programs allow text to overlay graphics, rotated text, and other elements which do not convert into HTML easily. The individual elements will be merged to become rasterized (pixel-based) graphics, which take longer to download than plain text. For this reason, a document should use a simple, column or table-based layout with text overlaying solid color areas.

Text size

If a layout is submitted as a rasterized graphic, the text will not be as clear as regular HTML text. This is because the text will be at the same resolution (72 dpi or dots per inch) as a web graphic. In order to avoid text which is difficult to read, text should be kept at a minimum size of 9 points.

Summary

Preparing Documents for PDF Format

When using ANY program to generate documents for PDF
- Include special fonts along with your document when delivered
- Use True Type fonts when possible
- Use Arial, Helvetica, Times, and Times New Roman for body text, and no more than two additional fonts for accents and headers
- Margins should be set to a minimum of .5 inches

When using Microsoft Word to generate documents for PDF
- When using section breaks, do not allow the margins to reset at less than .5 inches
- If a document requires different page orientations (Portrait/Landscape) provide a separate file for each orientation
- If accurate page numbering is required, insert Page Breaks between each page.
- Do not allow headers or footers to come less than .5 inches to the edge of the page
- Do not use text boxes

When using Microsoft Excel to generate documents for PDF
- Avoid large custom page sizes, or large spreadsheets which will cause the text to become too small when scaled to print.

When using Microsoft Word to generate documents for PDF
- Refer to *When using ANY program to generate documents for PDF* above

Preparing Documents for HTML Format

When using ANY program to prepare documents for HTML format
- Use Times, Arial, or Helvetica fonts in documents whenever possible
- Use fonts in the following sizes: 9 pt., 10pt., 12pt., 14pt.,18pt., 24pt.
- Provide graphics in GIF or JPG format if possible

When using Microsoft Word to generate documents for HTML
- Do not use text boxes

When using Microsoft Excel to generate documents for HTML
- If possible, create spreadsheets which will scroll vertically (narrow enough to fit on a computer screen)

When using layout programs (i.e Visio, Pagemaker, Powerpoint) to generate documents for HTML
- Complex layouts using overlayed text, etc. will be large files and will download slowly, so use simpler layouts with text over solid colors if possible
- Text which overlays images will be used as a graphic, so avoid small (below 9 pt.) text